Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 73(2) 2019 153–160

Outliers in Incomplete Multi-Response Experiments in Presence of Masking

Raju Kumar and L.M. Bhar
ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Received 18 December 2018; Revised 20 May 2019; Accepted 27 May 2019

SUMMARY

A method of identifying subset of outliers in presence of masking has been developed for incomplete Multi-Response design. Design is composed of two sets of experimental units. Different numbers of response variables are observed from these two sets. A Conditional Cook's Statistics in block design for incomplete multiresponse experiments has been developed for identification of outliers in presence of masking. The developed statistic has been illustrated with a real life data set. It has been shown that outliers in presence of masking can distort the overall conclusion from an experiment.

Keywords: Incomplete Multi-response experiments; Outlier; Masking; Cook-statistic.

1. INTRODUCTION

In statistics, there are many ways to obtain a data. Among them, conducting an experiment is one way to obtain the data. For obtaining a coherent analysis, it is necessary to detect an outlier(s) present in the data. The main reason for the importance of detection of outlier(s) is that potentially they have strong influence on the overall conclusion from an experiment. Therefore, it is important to detect and handle the outlier(s) efficiently. Outlier(s) in multi-response experiments whether complete or incomplete is/ are also likely to occur for many reasons. In any experiment, if an experimental plot is heavily infested with pests, disease, rodent and/or weeds then all the responses observed from that plot may be outlier(s). Again, if the data set contains multiple outliers or influential observation, which is more common in case of data sets, the problem of identifying such observations becomes more difficult due to masking effect. Masking occurs when one outlier is undetected because of the presence of other outliers. Cook (1977) introduced the well known Cook's distance for the identification of influential observations in linear regression. Cook and Weisberg (1982), Atkinson (1985) and Chatterjee and Hadi (1988) suggested a number of influential measures which are usually pay attention to detecting individually influential observation. Davies and Gather (1993), Hadi and Simonoff (1993) and Hadi (1994) developed some test procedure for detection of multiple outliers that are free from masking effect. A significant work in terms of identification of influential subsets in presence of masking in linear regression has been done by Pena and Yohi (1995). Lawrance (1995) suggested a conditional cook's distance to measure the influence of observations conditional on the prior removal of other cases. This masking problem has been dealt with in great detail in linear regression but may not get much attention in the context of experimental design. Bhar and Gupta (2001) developed a Cook-statistics for identification of outliers in design experiments when our interest is in estimation of some set of treatment contrast. Bhar *et. al.* (2013) proposed a method to identify outliers in presence of masking in a designed experiment. They formed a influential matrix in which elements of this matrix are derived from Cook-distance. In the present investigation, we developed a method to identify outliers in incomplete multiresponse design in

Corresponding author: Raju Kumar
E-mail address: raju.kumar@icar.gov.in

presence of masking when interest is in the estimation of treatment contrast.

2. INCOMPLETE MULTIRESPONSE DESIGN

Consider an experimental design in which p $(= p_1+p_2)$ responses are observed from $n=(n_1+n_2)$ experimental units. Let the $p_1(<p)$ response variables are observed from n_1 experimental units and $p_2(<p)$ response variables are observed from all $(n_1+n_2=n)$ experimental units. Thus, there are n_2 experimental units from where only p_2 observations are recorded. For first p_1 response variables, model can be written as

$$\mathbf{Y}_i = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\xi}_i,\; i=1, 2, \ldots, p_1, \tag{1}$$

where $\mathbf{y}_i$ is a $n_1 \times 1$ vector of observations on the i^{th} response variable, $\mathbf{X}_1$ is a matrix of known functions of the setting of coded variables, $\boldsymbol{\beta}_1$ is a $m_1 \times 1$ vector of unknown parameters and $\boldsymbol{\xi}_i$ is a $n_1 \times 1$ vector of random errors associated with the i^{th} response variable ($i=1, 2, \ldots, p_1$).

Model for p_2 response variables from n_1 experimental units can be written as

$$\mathbf{Y}_j = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\xi}_j,\; j = p_1+1, p_1+2, \ldots, p_1+p_2, \tag{2}$$

where $\mathbf{y}_j$ is a $n_1 \times 1$ vector of observations on the j^{th} response variable, $\mathbf{X}_1$ is a matrix of known functions of the setting of coded variables, $\boldsymbol{\beta}_1$ is a $m_1 \times 1$ vector of unknown parameters and $\boldsymbol{\xi}_j$ is a $n_1 \times 1$ vector of random errors associated with the j^{th} response variable ($j=p_1+1, p_1+2, \ldots, p_1+p_2$,).

Model for p_2 response variables from n_2 experimental units can be written as

$$\mathbf{Y}_k = \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\xi}_k,\; k = p_1+1, p_1+2, \ldots, p_1+p_2, \tag{3}$$

where $\mathbf{y}_k$ is a $n_2 \times 1$ vector of observations on the k^{th} response variable, $\mathbf{X}_2$ is a $n_2 \times m_2$ matrix of known functions of the setting of coded variables, $\boldsymbol{\beta}_2$ is a $m_2 \times 1$ vector of unknown parameters and $\boldsymbol{\xi}_k$ is a $n_2 \times 1$ vector of random errors associated with the k^{th} response variable ($k = p_1+1, p_1+2, \ldots, p_1+p_2$).

Now combining and roll down the models (1), (2) and (3) we have

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi}, \tag{4}$$

where

$$\begin{aligned}\mathbf{Y} &= \begin{bmatrix}\mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_{p_1} & \mathbf{y}_{p1+1} & \mathbf{y}_{p1+2} & \cdots & \mathbf{y}_{p_1+p_2}\end{bmatrix}' \\ &= \begin{bmatrix}\mathbf{Y}_1^* & \mathbf{Y}_2^* & \mathbf{Y}_2^{**}\end{bmatrix}' = \begin{bmatrix}\mathbf{Y}_1 & \mathbf{Y}_2\end{bmatrix}'\end{aligned} \tag{5}$$

$\mathbf{Y}_1^*$ is $(n_1p_1 \times 1)$ vector of observations on p_1 response variables, $\mathbf{Y}_2^*$ is $(n_1p_2 \times 1)$ vector of observations on p_2 response variables corresponding to first n_1 experimental units and $\mathbf{Y}_2^{**}$ is $(n_2p_2 \times 1)$ vector of observations on p_2 response variables for remaining n_2 experimental units. Combining $\mathbf{Y}_1^*$ and $\mathbf{Y}_2^*$, we get $\mathbf{Y}_1$ a n_1p vector of observations on p response variables and $\mathbf{Y}_2^{**} = \mathbf{Y}_2$. So $\mathbf{Y}$ is a $(n_1p_1 + n_1p_2 + n_2p_2)$ x1 vector of the observations on all $(p_1 + p_2)$ response variables and $\boldsymbol{\xi}$ is a $(n_1p_1 + n_1p_2 + n_2p_2)$ x1 vector of errors.

The design matrix $\mathbf{X}$ can be written as

$$\mathbf{X} = \begin{bmatrix}\mathbf{I}_{p_1} \otimes \mathbf{X}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p_2} \otimes \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{p_2} \otimes \mathbf{X}_2\end{bmatrix} = \begin{bmatrix}\mathbf{I}_p \otimes \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p_2} \otimes \mathbf{X}_2\end{bmatrix} \tag{6}$$

In the same way, the parameter vector $\boldsymbol{\beta}$ can be written as

$$\boldsymbol{\beta} = \begin{bmatrix}(\mathbf{1}_{p_1} \otimes \boldsymbol{\beta}_1) & (\mathbf{1}_{p_2} \otimes \boldsymbol{\beta}_1) & (\mathbf{1}_{p_2} \otimes \boldsymbol{\beta}_2)\end{bmatrix}' = \begin{bmatrix}\boldsymbol{\beta}_1^* & \boldsymbol{\beta}_2^*\end{bmatrix}' \tag{7}$$

where $\boldsymbol{\beta}_1^* = \begin{bmatrix}(\mathbf{1}_{p_1} \otimes \boldsymbol{\beta}_1) & (\mathbf{1}_{p_2} \otimes \boldsymbol{\beta}_1)\end{bmatrix}'$ is a $m_1p \times 1$ vector of parameters for p response variables on n_1 experimental units and $\boldsymbol{\beta}_2^*$ is a $m_2p_2 \times 1$ vector of parameters for p_2 response variables on n_2 experimental units. Similarly $\boldsymbol{\xi}$ is partitioned as

$$\boldsymbol{\xi} = \begin{bmatrix}\boldsymbol{\xi}_1 & \boldsymbol{\xi}_2 & \cdots & \boldsymbol{\xi}_{p_1} & \cdots & \boldsymbol{\xi}_{p_1+p_2}\end{bmatrix}' = \begin{bmatrix}\boldsymbol{\xi}_1^* & \boldsymbol{\xi}_2^*\end{bmatrix}' \tag{8}$$

where $\boldsymbol{\xi}_1^*$ is a $n_1p \times 1$ vector of errors for first p response variables on n_1 experimental units, $\boldsymbol{\xi}_2^*$ is a $n_2p_2 \times 1$ vector of errors for p_2 response variables on n_2 experimental units.

From the partitioned vectors and matrices from (5) to (7), it is clear that the design D is a incomplete multiresponse design having two sets S_1 and S_2 i.e. design D is composed of two sub design D_1 and D_2. Sub design D_1 consists of n_1 experimental units from which all p responses are observed where as D_2 is a sub design of n_2 experimental units from which only p_2 response variables are observed. Assuming that design **D** remains connected after deletion of any t observations, Kumar and Bhar (2017) have shown that the difference between the estimate of all contrasts of treatment effects for the whole design D can be written as

$$\mathbf{P}\,\hat{\boldsymbol{\theta}}_1^0 - \mathbf{P}\,\hat{\boldsymbol{\theta}}_{1(t)}^0 = \begin{pmatrix} \mathbf{P}_1\hat{\boldsymbol{\tau}}_1^* - \mathbf{P}_1\hat{\boldsymbol{\tau}}_{1(t)}^* \\ \mathbf{P}_2\hat{\boldsymbol{\tau}}_2^* - \mathbf{P}_2\hat{\boldsymbol{\tau}}_{2(t)}^* \end{pmatrix} = \begin{bmatrix} (\mathbf{I}_p \otimes \mathbf{P}^*\mathbf{C}_1^{*+}\mathbf{X}_{11}'\boldsymbol{\varphi}_1\mathbf{U}_1^*\left(\mathbf{U}_1^{*\prime}\mathbf{V}_1\mathbf{U}_1^*\right)^{-1}\mathbf{U}_1^{*\prime}\mathbf{V}_1)\mathbf{Y}_1 \\ (\mathbf{I}_{p_2} \otimes \mathbf{P}^*\mathbf{C}_2^{*+}\mathbf{X}_{22}'\boldsymbol{\varphi}_2\mathbf{U}_2^*\left(\mathbf{U}_2^{*\prime}\mathbf{V}_2\mathbf{U}_2^*\right)^{-1}\mathbf{U}_2^{*\prime}\mathbf{V}_2)\mathbf{Y}_2 \end{bmatrix} \tag{9}$$

$\mathbf{P}\,\hat{\boldsymbol{\theta}}_1^0$ is the least square estimators of $\mathbf{P}\,\boldsymbol{\theta}_1^0$, $\boldsymbol{\theta}_1^0 = \begin{bmatrix}\tau_1^* & \tau_1^*\end{bmatrix}'$ is a vector parameters of parameter of interest where $\boldsymbol{\tau}_1^* = \mathbf{I}_p \otimes \boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2^* = \mathbf{I}_{p_2} \otimes \boldsymbol{\tau}_2$. $\mathbf{P}\,\hat{\boldsymbol{\theta}}_{1(t)}^0$ is the least square estimators of $\mathbf{P}\,\boldsymbol{\theta}_{1(t)}^0$ obtained after deleting suspected t outlying observations, $\mathbf{U}_1 = \mathbf{I}_p \otimes \mathbf{U}_1^*$, $\mathbf{U}_1^* = \left(\mathbf{u}_1 \;\; \mathbf{u}_2 \;\; \mathbf{u}_{t_1}\right)$, $\mathbf{u}_i$ is a n_1 component vector having j^{th} element as 1 (if j^{th} observation is an outlier) and rest are zero. i.e., $\mathbf{u}_i = (0,0,...,0,1(j^{th}),0,0,...,0)'$, $\forall j = 1,2,...,n_1$, $i = 1, 2, \ldots, t_1 (t_1 < n_1)$, $\mathbf{U}_2 = \mathbf{I}_p \otimes \mathbf{U}_2^*$, $\mathbf{u}_i$ is a n_2 component vector having j^{th} element as 1(if j^{th} observation is an outlier) and rest are zero, i.e., $\mathbf{u}_i = (0,0,...,0,1(j^{th}),0,0,...,0)'$, $\forall j = 1,2,...,n_2$, $i = 1, 2, \ldots, t_2 (t_2 < n_2)$, $\boldsymbol{\varphi}_1 = (\mathbf{I} - \mathbf{X}_{12}(\mathbf{X}_{12}'\mathbf{X}_{12})^{-}\mathbf{X}_{12}')$,

$\mathbf{V}_1 = \Sigma_{11}(\boldsymbol{\varphi}_1 - \boldsymbol{\varphi}_1\mathbf{X}_{11}\mathbf{C}_1^{*-}\mathbf{X}_{11}'\boldsymbol{\varphi}_1)$ and $\mathbf{C}_1^{*+} / \mathbf{C}_2^{*+}$ is Moore-Penrose inverse of C-matrix for uni-response experiment under mean shift model [see Bhar & Ojha, 2014].

3. CONDITIONAL COOK'S STATISTICS IN BLOCK DESIGN FOR INCOMPLETE MULTIRESPONSE EXPERIMENTS

To develop a measure for assessing the masking effect in incomplete multiresponse design, we have to calculate conditional cook's distance to measure the influence of observations conditional on the prior removal of other. We, therefore, first develop this statistic for the case when our interest in estimation of some function of treatment effects. If a design D is an incomplete multiresponse design having two sets S_1 and S_2 i.e. design D is composed of two sub design D_1 and D_2 then the possible cases of occurrence of two outliers are, (i) Two outlier vectors are present in Set S_1 or two outlier vectors are present in of set S_2. (ii) Out of two outliers, one outlier vector is present in Set S_1 and another is present in set S_2. We will develop cook's distance for two cases to measure the masking effect.

Case–I: Two outliers occur in set-1 i.e. S_1 consisting of n_1 experimental units.

In the mean-shift outlier model, we defined $\mathbf{U}_1^* = \left(\mathbf{u}_1 \quad \mathbf{u}_2\right)$.

The difference between the estimates of all contrasts of treatment effects for the whole design ***D*** when two outliers occur in design $\mathbf{D}_1$ from equation (9) can be written as

$$\mathbf{P}\,\hat{\boldsymbol{\theta}}_1^0 - \mathbf{P}\,\hat{\boldsymbol{\theta}}_{1(ij)}^0 = \begin{bmatrix} (\mathbf{I}_p \otimes \mathbf{P}^*\mathbf{C}_1^{*+}\mathbf{X}_{11}'\boldsymbol{\varphi}_1\mathbf{U}_1^*\left(\mathbf{U}_1^{*\prime}\mathbf{V}_1\mathbf{U}_1^*\right)^{-1}\mathbf{U}_1^{*\prime}\mathbf{V}_1)\mathbf{Y}_1 \\ \mathbf{0} \end{bmatrix} \tag{10}$$

Now

$$\mathbf{U}_1^{*\prime}\mathbf{V}_1\mathbf{U}_1^* = \begin{bmatrix} \mathbf{u}_1'\mathbf{v}_1 \\ \mathbf{u}_2'\mathbf{v}_1 \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{u}_1'\sum_{11}(\boldsymbol{\varphi}_1 - \boldsymbol{\varphi}_1\mathbf{X}_{11}\mathbf{C}_1^*\mathbf{X}_{11}'\boldsymbol{\varphi}_1)\mathbf{u}_1 & \mathbf{u}_1'\sum_{11}(\boldsymbol{\varphi}_1 - \boldsymbol{\varphi}_1\mathbf{X}_{11}\mathbf{C}_1^*\mathbf{X}_{11}'\boldsymbol{\varphi}_1)\mathbf{u}_2 \\ \mathbf{u}_2'\sum_{11}(\boldsymbol{\varphi}_1 - \boldsymbol{\varphi}_1\mathbf{X}_{11}\mathbf{C}_1^*\mathbf{X}_{11}'\boldsymbol{\varphi}_1)\mathbf{u}_1 & \mathbf{u}_2'\sum_{11}(\boldsymbol{\varphi}_1 - \boldsymbol{\varphi}_1\mathbf{X}_{11}\mathbf{C}_1^*\mathbf{X}_{11}'\boldsymbol{\varphi}_1)\mathbf{u}_2 \end{bmatrix}$$

$$= \begin{bmatrix} h_{ii} & h_{ij} \\ h_{ji} & h_{jj} \end{bmatrix} \text{ (say)}$$

Hence,

$$\mathbf{P}\,\hat{\boldsymbol{\theta}}_1^0 - \mathbf{P}\,\hat{\boldsymbol{\theta}}_{1(ij)}^0 = \begin{bmatrix} \left(\mathbf{I}_p \otimes \mathbf{P}^*\mathbf{C}_1^{*+}\mathbf{X}_{11}'\boldsymbol{\varphi}_1\mathbf{U}_1^*\dfrac{1}{\Delta}\begin{bmatrix}(h_{jj}\boldsymbol{\omega}_1 - h_{ij}\boldsymbol{\omega}_2) \\ (h_{ii}\boldsymbol{\omega}_2 - h_{ji}\boldsymbol{\omega}_1)\end{bmatrix}\right)\mathbf{Y}_1 \\ \mathbf{0} \end{bmatrix}$$

where $\boldsymbol{\omega}_1 = \mathbf{u}_1'\mathbf{V}_1$ and $\boldsymbol{\omega}_2 = \mathbf{u}_2'\mathbf{V}_2$.

The difference between the estimates of all contrasts of treatment effects for the whole design **D** when single outliers (i^{th} observation) occur in design $\mathbf{D}_1$ (set S_1) can be written as

$$\mathbf{P}\,\hat{\boldsymbol{\theta}}_1^0 - \mathbf{P}\,\hat{\boldsymbol{\theta}}_{1(i)}^0 = \begin{bmatrix} (\mathbf{I}_p \otimes \mathbf{P}^*\mathbf{C}_1^{*+}\mathbf{X}_{11}'\boldsymbol{\varphi}_1\mathbf{u}_1\left(\mathbf{u}_1'\mathbf{V}_1\mathbf{u}_1\right)^{-1}\mathbf{u}_1'\mathbf{V}_1)\mathbf{Y}_1 \\ \mathbf{0} \end{bmatrix} \tag{11}$$

$$\text{or, } \mathbf{P}\,\hat{\boldsymbol{\theta}}_1^0 - \mathbf{P}\,\hat{\boldsymbol{\theta}}_{1(i)}^0 = \begin{bmatrix} \left(\mathbf{I}_p \otimes \mathbf{P}^*\mathbf{C}_1^{*+}\mathbf{X}_{11}'\boldsymbol{\varphi}_1\mathbf{u}_1\dfrac{\boldsymbol{\omega}_1}{h_{ii}}\right)\mathbf{Y}_1 \\ \mathbf{0} \end{bmatrix}$$

The difference between $\mathbf{P}\,\hat{\boldsymbol{\theta}}^0_{1(ij)}$ and $\mathbf{P}\,\hat{\boldsymbol{\theta}}^0_{1(i)}$

$$\mathbf{P}\,\hat{\boldsymbol{\theta}}^0_{1(ij)} - \mathbf{P}\,\hat{\boldsymbol{\theta}}^0_{1(i)} = \left[\mathbf{I}_p \otimes \mathbf{P}^*\mathbf{C}_1^{*+}\mathbf{X}'_{11}\boldsymbol{\varphi}_1\mathbf{u}_1\frac{\boldsymbol{\omega}_1}{h_{ii}}\right]\mathbf{Y}_1 - \begin{bmatrix}\mathbf{I}_p \otimes \mathbf{P}^*\mathbf{C}_1^{*+}\mathbf{X}'_{11}\boldsymbol{\varphi}_1\mathbf{U}_1^*\dfrac{1}{\Delta}\begin{bmatrix}(h_{jj}\boldsymbol{\omega}_1 - h_{ij}\boldsymbol{\omega}_2)\\(h_{ii}\boldsymbol{\omega}_2 - h_{ji}\boldsymbol{\omega}_1)\end{bmatrix}\\ \mathbf{0}\end{bmatrix}\mathbf{Y}_1$$

Hence,

$$\mathbf{P}\,\hat{\boldsymbol{\theta}}^0_{1(ij)} - \mathbf{P}\,\hat{\boldsymbol{\theta}}^0_{1(i)} = \begin{bmatrix}\left(\mathbf{I}_p \otimes \mathbf{P}^*\mathbf{C}_1^{*+}\mathbf{X}'_{11}\boldsymbol{\varphi}_1\boldsymbol{\eta}\right)\mathbf{Y}_1\\ \mathbf{0}\end{bmatrix}, \quad (12)$$

where

$$\boldsymbol{\eta} = \left(0 \;\ldots\; \frac{1}{\Delta}(h_{ii}\boldsymbol{\omega}_2 - h_{ji}\boldsymbol{\omega}_1) \;\ldots\; \left(\frac{1}{h_{ii}} - \frac{1}{\Delta}h_{jj}\right)\grave{\mathbf{u}}_1 - h_{ij}\grave{\mathbf{u}}_2 \;\ldots\; 0\right)'$$

Now,

$$V(\mathbf{P}\hat{\boldsymbol{\theta}}^0_{1(i)}) = \left(\mathbf{P}\mathbf{C}_{(i)}\mathbf{P}'\right)^{-1} = \begin{pmatrix}\mathbf{P}_1\mathbf{C}_{1(i)}\mathbf{P}_1' & \mathbf{0}\\ \mathbf{0} & \mathbf{P}_2\mathbf{C}_2\mathbf{P}_2'\end{pmatrix}^{-1}$$

$$= \begin{bmatrix}\sum_{11}\otimes(\mathbf{P}^*\mathbf{C}^*_{1(i)}\;\mathbf{P}^{*'})^{-1} & \mathbf{0}\\ \mathbf{0} & \sum_{22}\otimes(\mathbf{P}^*\mathbf{C}^*_2\;\mathbf{P}^{*'})^{-1}\end{bmatrix}$$

and

$$\mathbf{C}^*_{1(i)} = \mathbf{C}^*_1 - \mathbf{X}_{11}\phi_1\mathbf{U}^*_1(\mathbf{U}^{*'}_1\phi_1\mathbf{U}^*_1)^{-1}\mathbf{U}^*_1\phi_1\mathbf{X}'_{11}.$$

Cook's distance for j^{th} observation (occur in $\mathbf{D}_1$) after the deletion of i^{th} observation (occur in $\mathbf{D}_1$) is defined as

$$C_{j(i)} = \frac{(\mathbf{P}\hat{\boldsymbol{\theta}}^0_{1(i)} - \mathbf{P}\hat{\boldsymbol{\theta}}^0_{1(ij)})'[V(\mathbf{P}\hat{\boldsymbol{\theta}}^0_{1(i)})]^{-1}(\mathbf{P}\hat{\boldsymbol{\theta}}^0_{1(i)} - \mathbf{P}\hat{\boldsymbol{\theta}}^0_{1(ij)})}{Rank[V(\mathbf{P}\hat{\boldsymbol{\theta}}^0_{1(i)})]} \quad (13)$$

Now substituting the value of $\mathbf{P}\,\hat{\boldsymbol{\theta}}^0_{1(ij)} - \mathbf{P}\,\hat{\boldsymbol{\theta}}^0_{1(i)}$ from (12), we get

$$C_{j(i)} = \frac{\mathbf{Y}_1'\left(\sum_{11}\otimes\boldsymbol{\eta}'\phi_1\mathbf{X}'_{11}\mathbf{C}^{*+}_{1(i)}\mathbf{X}_{11}\boldsymbol{\varphi}_1\boldsymbol{\eta}\right)\mathbf{Y}_1}{(p+p_2)(v-1)} \quad (14)$$

Cook's distance for single outlier vector (occur in $\mathbf{D}_1$)

$$\mathbf{P}\,\hat{\boldsymbol{\theta}}^0_1 - \mathbf{P}\,\hat{\boldsymbol{\theta}}^0_{1(i)} = \begin{bmatrix}\left(\mathbf{I}_p \otimes \mathbf{P}^*\mathbf{C}_1^{*+}\mathbf{X}'_{11}\boldsymbol{\varphi}_1\mathbf{u}_1\dfrac{\boldsymbol{\omega}_1}{h_{ii}}\right)\mathbf{Y}_1\\ \mathbf{0}\end{bmatrix}$$

Now, $\mathbf{I}_p \otimes \mathbf{P}^*\mathbf{C}_1^{*+}\mathbf{X}'_{11}\boldsymbol{\varphi}_1\mathbf{u}_1\dfrac{\omega_1}{h_{ii}} = \mathbf{I}_p \otimes \mathbf{P}^*\mathbf{C}_1^{*+}\mathbf{X}'_{11}\boldsymbol{\varphi}_1\varpi_1$,

where $\varpi_1 = \left[0 \;\ldots\; \dfrac{\boldsymbol{\omega}_1}{h_{ii}} \;\ldots\; 0\right]'$

Hence, Cook's distance for single outliers (i^{th} observation)

$$C_i = \frac{\mathbf{Y}_1'\left(\sum_{11}\otimes\varpi_1'\boldsymbol{\varphi}_1\mathbf{X}'_{11}\mathbf{C}_1^{*+}\mathbf{X}_{11}\boldsymbol{\varphi}_1\varpi_1\right)\mathbf{Y}_1}{(p+p_2)(v-1)} \quad (15)$$

Following Lawrance (1995), the masking factor $M_{j(i)}$ is defined as

$$M_{j(i)} = C_{j(i)}/C_i = \frac{\mathbf{Y}_1'\left(\sum_{11}\otimes\boldsymbol{\eta}'\phi_1\mathbf{X}'_{11}\mathbf{C}^{*+}_{1(i)}\mathbf{X}_{11}\boldsymbol{\varphi}_1\boldsymbol{\eta}\right)\mathbf{Y}_1}{\mathbf{Y}_1'\left(\sum_{11}\otimes\varpi_1'\boldsymbol{\varphi}_1\mathbf{X}'_{11}\mathbf{C}_1^{*+}\mathbf{X}_{11}\boldsymbol{\varphi}_1\varpi_1\right)\mathbf{Y}_1} \quad (16)$$

Similarly, the masking factor $M_{j(i)}$ for j^{th} observation (occur in $\mathbf{D}_2$) after the deletion of i^{th} observation (occur in $\mathbf{D}_2$) is defined as

$$M_{j(i)} = \frac{\mathbf{Y}_2'\left(\sum_{22}\otimes\boldsymbol{\eta}_0'\phi_2\mathbf{X}'_{21}\mathbf{C}^{*+}_{1(i)}\mathbf{X}_{21}\boldsymbol{\varphi}_1\boldsymbol{\eta}_0\right)\mathbf{Y}_2}{\mathbf{Y}_2'\left(\sum_{22}\otimes\varpi_{02}'\ddot{o}_2\mathrm{X}'_{21}\mathrm{C}_2^{*+}\mathrm{X}_{21}\ddot{o}_2\varpi_{02}\right)\mathbf{Y}_2}, \quad (17)$$

where

$$\boldsymbol{\eta}_0 = \left(0 \;\ldots\; \frac{1}{\Delta^*}(h^*_{ii}\boldsymbol{\omega}^*_1 - h^*_{ij}\boldsymbol{\omega}^*_1) \;\ldots\; \left(\frac{1}{h^*_{ii}} - \frac{1}{\Delta^*}h_{jj}\right)\boldsymbol{\omega}^*_1 - h^*_{ij}\boldsymbol{\omega}^*_2 \;\ldots\; 0\right)',$$

$\Delta^* = h^*_{ii}h^*_{jj} - h^*_{ji}h^*_{ij}$, $\boldsymbol{\omega}^*_1 = \mathbf{u}_1'\mathbf{V}_2$, $\omega^*_2 = \mathbf{u}_2'\mathbf{V}_2$,

$h^*_{ii} = \mathbf{u}_1'\sum_{22}(\boldsymbol{\varphi}_2 - \boldsymbol{\varphi}_2\mathbf{X}_{21}\mathbf{C}^*_2\mathbf{X}'_{21}\boldsymbol{\varphi}_2)\mathbf{u}_1$,

$h^*_{ij} = \mathbf{u}_1'\sum_{22}(\boldsymbol{\varphi}_2 - \boldsymbol{\varphi}_2\mathbf{X}_{21}\mathbf{C}^*_2\mathbf{X}'_{21}\boldsymbol{\varphi}_2)\mathbf{u}_2$,

$h^*_{ji} = \mathbf{u}_2'\sum_{22}(\boldsymbol{\varphi}_2 - \boldsymbol{\varphi}_2\mathbf{X}_{21}\mathbf{C}^*_2\mathbf{X}'_{21}\boldsymbol{\varphi}_2)\mathbf{u}_1$ and

$h^*_{jj} = \mathbf{u}_2'\sum_{22}(\boldsymbol{\varphi}_2 - \boldsymbol{\varphi}_2\mathbf{X}_{21}\mathbf{C}^*_2\mathbf{X}'_{21}\boldsymbol{\varphi}_2)\mathbf{u}_2$.

Here, if $M_{j(i)} > 1$, then we conclude that i^{th} observation is masked by the j^{th} outlier and both the observation are termed as outliers.

Case-II. When one outlier occurs in $\mathbf{D}_1$ (set S_1) and other in $\mathbf{D}_2$ (set S_2) then in mean shift model, $\mathbf{U}^*_1 = (\mathbf{u}_1)$ and $\mathbf{U}^*_2 = (\mathbf{u}_2)$.

$$\mathbf{P}\,\hat{\boldsymbol{\theta}}_1^0 - \mathbf{P}\,\hat{\boldsymbol{\theta}}_{1(ij)}^0 = \begin{bmatrix} \left[\mathbf{I}_p \otimes \mathbf{P}^*\mathbf{C}_1^{*+}\mathbf{X}_{11}'\boldsymbol{\varphi}_1\mathbf{u}_1\left(\mathbf{u}_1'\Sigma_{11}(\boldsymbol{\varphi}_1 - \boldsymbol{\varphi}_1\mathbf{X}_{11}\mathbf{C}_1^{*-}\mathbf{X}_{11}'\boldsymbol{\varphi}_1)\mathbf{u}_1\right)^{-1}\mathbf{u}_1'\Sigma_{11}(\boldsymbol{\varphi}_1 - \boldsymbol{\varphi}_1\mathbf{X}_{11}\mathbf{C}_1^{*-}\mathbf{X}_{11}'\boldsymbol{\varphi}_1)\right]\mathbf{Y}_1 \\ \left[\mathbf{I}_{p_2} \otimes \mathbf{P}^*\mathbf{C}_2^{*+}\mathbf{X}_{21}'\boldsymbol{\varphi}_2\mathbf{u}_2\left(\mathbf{u}_2'\Sigma_{22}(\boldsymbol{\varphi}_2 - \boldsymbol{\varphi}_2\mathbf{X}_{21}\mathbf{C}_2^{*-}\mathbf{X}_{21}'\boldsymbol{\varphi}_2)\mathbf{u}_2\right)^{-1}\mathbf{u}_2'\Sigma_{22}(\boldsymbol{\varphi}_2 - \boldsymbol{\varphi}_2\mathbf{X}_{21}\mathbf{C}_2^{*-}\mathbf{X}_{21}'\boldsymbol{\varphi}_2)\right]\mathbf{Y}_2 \end{bmatrix}$$

$$= \begin{bmatrix} \left(\mathbf{I}_p \otimes \mathbf{P}^*\mathbf{C}_1^{*+}\mathbf{X}_{11}'\boldsymbol{\varphi}_1\mathbf{u}_1\dfrac{\boldsymbol{\omega}_\mathbf{i}}{h_\mathbf{ii}}\right)\mathbf{Y}_1 \\ \left(\mathbf{I}_{p_2} \otimes \mathbf{P}^*\mathbf{C}_2^{*+}\mathbf{X}_{21}'\boldsymbol{\varphi}_2\mathbf{u}_2\dfrac{\boldsymbol{\omega}_j}{h_{jj}}\right)\mathbf{Y}_2 \end{bmatrix} \tag{18}$$

The difference between the estimates of all contrasts of treatment effects for the whole design D when single outliers (i^{th} observation) occur in design $\mathbf{D}_1$ can be written as

$$\mathbf{P}\,\hat{\boldsymbol{\theta}}_1^0 - \mathbf{P}\,\hat{\boldsymbol{\theta}}_{1(i)}^0 = \begin{bmatrix} \left(\mathbf{I}_p \otimes \mathbf{P}^*\mathbf{C}_1^{*+}\mathbf{X}_{11}'\boldsymbol{\varphi}_1\mathbf{u}_1\dfrac{\boldsymbol{\omega}_1}{h_{ii}}\right)\mathbf{Y}_1 \\ \mathbf{0} \end{bmatrix} \tag{19}$$

The difference between $\mathbf{P}\,\hat{\boldsymbol{\theta}}_{1(ij)}^0$ and $\mathbf{P}\,\hat{\boldsymbol{\theta}}_{1(i)}^0$

$$\mathbf{P}\,\hat{\boldsymbol{\theta}}_{1(ij)}^0 - \mathbf{P}\,\hat{\boldsymbol{\theta}}_{1(i)}^0 = \begin{bmatrix} \mathbf{0} \\ \left(\mathbf{0} - \left(\mathbf{I}_{p_2} \otimes \mathbf{P}^*\mathbf{C}_2^{*+}\mathbf{X}_{21}'\boldsymbol{\varphi}_2\varpi_2\right)\right)\mathbf{Y}_2 \end{bmatrix},$$

where $\varpi_2 = \mathbf{u}_2\dfrac{\boldsymbol{\omega}_j}{h_{jj}}$)

Now,

$$V(\mathbf{P}\hat{\boldsymbol{\theta}}_{1(i)}^0) = \left(\mathbf{P}\mathbf{C}_{(i)}\mathbf{P}'\right)^{-1}\begin{pmatrix} \mathbf{P}_1\mathbf{C}_{1(i)}\mathbf{P}_1' & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_2\mathbf{C}_2\mathbf{P}_2' \end{pmatrix}^{-1} = \begin{bmatrix} \sum_{11}\otimes(\mathbf{P}^*\mathbf{C}_{1(i)}^*\ \mathbf{P}^{*'})^{-1} & \mathbf{0} \\ \mathbf{0} & \sum_{22}\otimes(\mathbf{P}^*\mathbf{C}_2^*\ \mathbf{P}^{*'})^{-1} \end{bmatrix}$$

and $\mathbf{C}_{1(i)}^* = \mathbf{C}_1^* - \mathbf{X}_{11}\phi_\mathbf{1}\mathbf{U}_\mathbf{1}^*(\mathbf{U}_\mathbf{1}^{*'}\phi_\mathbf{1}\mathbf{U}_\mathbf{1}^*)^{-1}\mathbf{U}_\mathbf{1}^*\phi_\mathbf{1}\mathbf{X}_\mathbf{11}'$

Cook's distance for j^{th} observation after the deletion of i^{th} observation is

$$C_{j(i)} = \frac{\mathbf{Y}_2'\left(\sum_{22}\otimes\varpi_2'\phi_\mathbf{2}\mathbf{X}_\mathbf{21}'\mathbf{C}_\mathbf{2}^{*+}\mathbf{X}_\mathbf{21}\boldsymbol{\varphi}_\mathbf{2}\varpi_2\right)\mathbf{Y}_2}{(p+p_2)(v-1)} \tag{20}$$

Cook's distance for single outliers (i^{th} observation occurs in Design $\mathbf{D}_1$)

$$C_i = \frac{\mathbf{Y}_1'\left(\sum_\mathbf{11}\otimes\varpi_1'\boldsymbol{\varphi}_\mathbf{1}\mathbf{X}_\mathbf{11}'\mathbf{C}_\mathbf{1}^{*+}\mathbf{X}_\mathbf{11}\boldsymbol{\varphi}_\mathbf{1}\varpi_1\right)\mathbf{Y}_1}{(p+p_2)(v-1)}. \tag{21}$$

Following Lawrance (1995), the masking factor $M_{j(i)}$ is defined as

$$M_{j(i)} = C_{j(i)}/C_i = \frac{\mathbf{Y}_2'\left(\sum_{22}\otimes\varpi_2'\phi_\mathbf{2}\mathbf{X}_\mathbf{21}'\mathbf{C}_\mathbf{2}^{*+}\mathbf{X}_\mathbf{21}\boldsymbol{\varphi}_\mathbf{2}\varpi_2\right)\mathbf{Y}_2}{\mathbf{Y}_1'\left(\sum_\mathbf{11}\otimes\varpi_1'\boldsymbol{\varphi}_\mathbf{1}\mathbf{X}_\mathbf{11}'\mathbf{C}_\mathbf{1}^{*+}\mathbf{X}_\mathbf{11}\boldsymbol{\varphi}_\mathbf{1}\varpi_1\right)\mathbf{Y}_1}. \tag{22}$$

Similarly, the masking factor $M_{j(i)}$ when i^{th} observation vector occurs in Design $\mathbf{D}_2$ and j^{th} observation vector occurs in Design $\mathbf{D}_1$

$$M_{j(i)} = \frac{\mathbf{Y}_1'\left(\sum_\mathbf{11}\otimes\varpi_1'\boldsymbol{\varphi}_\mathbf{1}\mathbf{X}_\mathbf{11}'\mathbf{C}_\mathbf{1}^{*+}\mathbf{X}_\mathbf{11}\boldsymbol{\varphi}_\mathbf{1}\varpi_1\right)\mathbf{Y}_1}{\mathbf{Y}_2'\left(\sum_{22}\otimes\varpi_2'\boldsymbol{\varphi}_\mathbf{2}\mathbf{X}_\mathbf{21}'\mathbf{C}_\mathbf{2}^{*+}\mathbf{X}_\mathbf{21}\boldsymbol{\varphi}_\mathbf{2}\varpi_2\right)\mathbf{Y}_2} \tag{23}$$

Here, if $M_{j(i)} > 1$, then we conclude that i^{th} observation vector is masked by the j^{th} outlier vector and both the observation vectors are termed as outliers.

4. ILLUSTRATION

An experiment with 9 treatments was conducted during the year 2011 in Uttar Pradesh to study the effect of integrated nutrient management on growth and yield of paddy. The experiment was laid out in Resolvable BIBD with 9 treatments ($v=9$, $b=12$, $r=4$, $k=3$, $\lambda=1$). The data on following 9 characters were observed: x_1 = plant height at harvest (cms), x_2 = dry matter (DM) accumulation at 90 days after sowing (DAS), x_3 = leaf area index (LAI) at 75 DAS, x_4 = number of spikes/sq. cm, x_5 = number of grains per spike, x_6 = test weight (gms), x_7 = grain yield (q/ha), x_8 = straw yield (q/ha) and x_9 = harvest index (%). Originally the experiment was laid out in Randomized Complete Block (RCB) design and for complete multi-response case. However, for illustration purpose, the layout of the experiment was made for a BIBD and for incomplete responses by deleting some observations appropriately. There are two total 36 experimental units, *i.e.*, n = 36 are divided in to two subsets each having 18 units. Treatments are distributed in to two sets in the following way:

Subset-1			Subset-2		
1	2	3	1	2	3
4	5	6	6	4	5
7	8	9	8	9	7
1	4	7	1	2	3
2	5	8	5	6	4
3	6	9	9	7	8

There are 9 response variables. Observations are taken for all 9 response variables from 18 experimental units and from remaining 18 units only 6 response variables are observed. Thus, we get $n_1=18$, $n_2=18$, $p=9$, $p_1=3$ and $p_2=6$. The data corresponding to each response variable is given in Table 1.

Table 1. Data corresponding to each response variable

S.N.	Trt.	Blk	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
1	1	1	100	708.9	3.2	314.9	32.4	38.1	23.9	38.5	34.3
2	2	1	118.7	714.7	4.1	298.2	35.2	45.7	28.0	31.3	33.3
3	3	1	110.7	730.4	3.1	326.6	34.2	48.5	21.9	32.1	23.4
4	4	2	99.1	769.2	3.6	341.5	36.3	38.6	25.4	40.8	36.3
5	5	2	109.8	750	3.7	324.8	33.3	43.9	23.6	39.1	36.9
6	6	2	100	719.6	3.1	316.5	30.4	40.1	25.4	41.7	38.7
7	7	3	93.8	708.3	3.1	303	31.4	37.7	23.1	37.5	34.1
8	8	3	98.2	719.6	3.3	296.3	30.4	42.9	23.6	40.0	34.2
9	9	3	97.3	705.9	3.1	324.8	34.2	39.2	22.8	37.3	35.1
10	1	4	93.2	701.3	3.1	298.4	31.6	37.7	22.7	46.9	34.1
11	4	4	97.3	711.8	3.4	313.8	32.3	42.9	21.2	45.2	32.4
12	7	4	95.5	730.4	3.3	298.2	35.2	40.1	22.8	44.3	35.1
13	2	5	95.3	714.9	3.1	309.4	42.8	37.8	23.5	38.1	34.2
14	5	5	98.2	750	3.4	307.3	43.8	36.4	23.6	35.6	32.4
15	8	5	95.5	739.2	3.1	313.8	39.1	39.2	21.2	36.5	36.9
16	3	6	92.1	690.4	2.9	293.4	30.4	37.3	22.3	36.8	33.7
17	6	6	97.3	750	3.1	296.3	27.6	40.1	21.0	37.3	34.2
18	9	6	99.1	688.2	3.3	286.2	30.4	43.9	22.8	35.6	32.4
19	1	7	-	-	-	289.3	30.5	37.3	29.8	36.3	33.6
20	2	7	-	-	-	286.2	31.4	42.9	28.1	39.1	31.5
21	3	7	-	-	-	294.5	33.3	39.2	25.4	37.3	32.4
22	6	8	-	-	-	303.9	31.5	37.7	22.8	37.4	33.8
23	4	8	-	-	-	307.3	29.5	42.9	23.6	36.5	35.1
24	5	8	-	-	-	316.5	34.2	44.8	22.8	39.1	36.9
25	8	9	-	-	-	332.6	35.2	38.3	25.1	40.3	34.4
26	9	9	-	-	-	334.9	34.2	38.3	24.5	37.3	33.3
27	7	9	-	-	-	334.9	36.1	42.9	25.4	36.5	34.2
28	1	10	-	-	-	328.3	34.2	38.2	23.	39.7	34.4
29	2	10	-	-	-	327.5	38.1	39.2	21.	41.7	35.1
30	3	10	-	-	-	345	33.3	38.3	21.	39.1	30.6
31	5	11	-	-	-	333	35.7	38.5	25.9	40.7	34.8
32	6	11	-	-	-	355	36.1	40.1	21.	37.3	32.4
33	4	11	-	-	-	336.7	34.2	38.3	24.5	33.9	33.3
34	9	12	-	-	-	324.6	34.2	37.9	24.5	39.3	34.4
35	7	12	-	-	-	326.6	33.3	39.2	25.4	39.1	31.5
36	8	12	-	-	-	324.8	30.4	40.1	21.1	40.8	36.9

In the present example, we have 9 treatments. Therefore, there will be 8 treatment contrasts for each of the response variables. Let this set of 8 treatment contrasts be given by a matrix $\mathbf{P}^*$, where $\mathbf{P}^*$ ={1 -1 0 0 0 0 0 0 0, 1 0 -1 0 0 0 0 0 0, 1 0 0 -1 0 0 0 0 0, 1 0 0 0 -1 0 0 0 0, 1 0 0 0 0 -1 0 0 0, 1 0 0 0 0 0 -1 0 0, 1 0 0 0 0 0 0 -1 0, 1 0 0 0 0 0 0 0 -1}. Combining these sets for all response variables, we get $\mathbf{P}_1 = \mathbf{I}_9 \otimes \mathbf{P}^*$, $\mathbf{P}_2 = \mathbf{I}_6 \otimes \mathbf{P}^*$ and we define $\mathbf{P} = \begin{bmatrix} \mathbf{P}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_2 \end{bmatrix}$. We denote by $\boldsymbol{\theta}$ the vector of treatment effects. Then we want to test the significance of the treatment contrast $H_0 : \mathbf{P}'\boldsymbol{\theta} = 0$ against $H_1 : \mathbf{P}'\boldsymbol{\theta} \neq 0$.

Test statistics for testing the null hypothesis is given by (Nandi, 2007)

$$TC = (\mathbf{P}'\hat{\boldsymbol{\theta}})'[D(\mathbf{P}'\hat{\boldsymbol{\theta}})]^{-1}(\mathbf{P}'\hat{\boldsymbol{\theta}})$$

Test statistics given above is asymptotically χ^2 distribution with one degree freedom. Therefore, reject the null hypothesis at $\alpha\%$ level of significance if $TC > \chi^2_{1-\alpha,1}$ and conclude that the treatment effects are significantly different. The calculated value of *TC* is found to be 14.31. Since this value is greater than the tabulated value of χ^2 at 1 degree of freedom, we reject the null hypothesis and conclude that there is a significance difference between the treatment 1 and treatment 2. Similarly, contrasts for other treatment comparisons are tested, and it was found that all treatment contrasts are statistically significant. Thus we conclude that overall treatment effects are significantly different.

We now apply the test-statistic as developed for detecting outliers (Kumar and Bhar (2017)), if any. We confined ourselves for the occurrence of a single outlier vector. Cook-statistics are calculated for all possible outlier observation vectors for both the sets. A program is written in SAS/IML to calculate these values. Calculated Cook-statistics are given in Table 2.

Table 2. Cook-statistics

Set 1				Set 2			
Obs. No	**Cook-statistic**	**Obs. No**	**Cook-statistic**	**Obs. No**	**Cook-statistic**	**Obs. No**	**Cook-statistic**
1	0.69058	10	0.31057	1	0.01165	10	0.01165
2	1.59405	**11**	**2.36571**	2	0.00469	11	0.00469
3	0.57604	12	0.59156	3	0.01089	12	0.01089
4	0.68347	13	0.27196	4	0.02396	13	0.01712
5	0.31863	14	0.24163	5	0.00111	14	0.02396
6	0.93688	15	0.17398	6	0.01712	15	0.00111
7	0.41874	16	0.18916	7	0.00162	16	0.0007
8	0.68174	17	0.65519	8	0.0007	17	0.00127
9	1.67455	18	0.2156	9	0.00127	18	0.00162

To determine the cut-off value, we worked out the inter-quartile range of this series of Cook-statistic values. This range is found to be $\tilde{S} = 0.4790$ for set 1 and $\tilde{S} = 0.1180$ for set 2. Then following Tukey (1977) the cut-off value is calculated and it is $7/2 \times \tilde{S} = 1.6768$ for Set 1 and $7/2 \times \tilde{S} = 0.3880$ for Set 2. Only one value of Cook-statistic exceeds this cut-off value. This is for observation number 11 in Set 1. Therefore, we conclude that the observation vector 11 is an outlier. This vector pertains to treatment number 4 in block 4.

We again conduct multivariate treatment contrast analysis after deleting this observation number 11 and calculated *TC* values. For example, this value is found out 9.46 for first contrast. This is again greater than calculated value. Thus, this treatments contrast remained as significant even after removing the outlying observation vector. However, there is change in the calculated value. *TC* value has been decreased significantly. Same thing is observed for other treatment contrast analysis. In present example, though for some treatment contrasts, the significant values are changed drastically, there is no change in overall conclusion. Thus the observation number 11, in spite of being outliers, does not have much influence in the conclusion. However, there might be groups of outliers that cannot be detected by using single outlier detection technique of Cook-statistics. We now apply the method developed in this paper. We calculate the conditional Cook-statistic and hence masking factor for all possible pairs of observations. For calculating the masking factor a program is written in SAS/IML. These values are presented in Table 3.

There are total 36 observations. Hence there would be ${}^{36}C_2 = 630$ pairs of combinations of observations, hence $M_{j(i)}$ values. In Table 3, we present some selected values of this statistic. As discussed earlier, if this value of $M_{j(i)}$ is greater than one, then we say that the ith observation is masked by j^{th} observation. We find that maximum value of this statistic is 2.60, which corresponds to observations no. 11 corresponding to treatment no. 4 in the 4th block and observation 1,

Table 3. Conditional Influence Statistics $M_{j(i)}$

S.N.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
J	4	12	16	23	11	32	10	21	22	4	19	13	21	1	26
I	1	1	2	12	8	12	8	28	32	7	5	18	33	11	24
$\mathbf{M}_{j(i)}$	0.43	0.37	0.09	0.48	0.99	0.64	0.58	0.72	0.47	0.62	0.50	0.27	0.4	**2.60**	0.11

which belongs to treatment no. 1 in the 1st block. Thus, the observation number 11 has masked the effect of observation number 1. The interesting point to be noted here is that though the observation number 11 was detected as outlier when we applied Cook-statistics for detecting a single outlier vector, yet observation vector number 1 was not. Its effect was masked by the observation number 11.

We re-conduct multivariate treatment contrast analysis after deleting these two observations number 11 and 1 and calculated *TC* values. The dramatic effect to note here that treatment contrast for first and ninth treatment became non-significant. Other treatment effects remained as significant. Thus, there is change in overall conclusion. Now, once observation number 1 and 11 are detected as outlier vectors, one would be interested to know the possible cause for the same. First thing, one should check if any error is committed during the time of data entry. That possibility is ruled out for present example. As, mentioned earlier, in field experiments, observations in some of the plots or blocks may come out to be abnormal due to uneven application of fertilizer and/or irrigation or other agronomic practices. The present example pertains to paddy crop. We observed two outliers vectors in which one outlier vector is masked by second outlier vector. This may be due to heavy/less irrigation in some plots.

REFERENCES

Atkinson, A.C. (1985). *Plots, transformation and regression*, Oxford University Press (pp292): Oxford.

Bhar, L. and Gupta, V.K. (2001). A useful statistic for studying outliers in experimental designs. *Sankhya* **B**, **63**, 338-350.

Bhar, L., Gupta, V.K. and Prasad, R. (2013). Detection of outliers in designed experiments in presence of masking. *Statistics and Application,* **11(1 & 2)**, 147-160.

Bhar, L. and Ojha, S. (2014). Outliers in multi-response experiments. *Commun. Statist. Theory Methods*. **43(13)**, 2782-2798.

Chatterjee, S. and Hadi, A.S. (1988). *Sensitivity Analysis in Linear Regression*. John Wiley: New York.

Cook, R.D. and Weisberg, S. (1982). *Residuals and influence in regression*, Chapman and Hall, London.

Cook, R.D. (1977). Detection of influential observation in linear regression. *Technometrics*, **19(1)**, 15-18.

Davies, L. and Gather, U. (1993). Identification of multiple outliers. *J. Amer. Statist.Assoc*., **88**, 782-801.

Hadi, A.S. (1994). A modification of a method for the detection of outliers in multivariate samples. *J. Roy. Statist. Soc.* B, **56**, 393-396.

Hadi, A.S. and Simonoff, J.S. (1993). Procedures for the identification of multiple outliers in linear models. *J. Amer. Statist. Assoc*., **88**, 1264-1272.

Kumar, R. and Bhar, L.M. (2017). Outliers in incomplete multiresponse experiments. *Commun. Statist. Theory Methods*, **46**(22**)**, 10426-10445.

Lawrance, A.J. (1995). Deletion influence and masking in regression**.** *J. Roy. Statist. Soc.* B, **57(1)**, 181-189.

Nandi, P.K. (2007). *Design and Analysis of Multi-response Experiments.* Ph. D. Thesis, IARI, New Delhi.

Pena, D. and Yohai, V.J. (1995). The detection of influential subsets in linear regression by using an influence matrix. *J. Roy. Statist. Soc.* B, **57**, 145-156.